**NLP Assignment-7**

1. **Explain the architecture of BERT.**

“BERT stands for Bidirectional Encoder Representations from Transformers. It is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context.

1. **Explain Masked Language Modeling (MLM)**

Masked language modeling (MLM), a self-supervised pretraining objective, is widely used in natural language processing for learning text representations. MLM trains a model to predict a random sample of input tokens that have been replaced by a [MASK] placeholder in a multi-class setting over the entire vocabulary.

1. **Explain Next Sentence Prediction (NSP)**

Next sentence prediction (NSP) is one-half of the training process behind the BERT model (the other being masked-language modeling — MLM). Where MLM teaches BERT to understand relationships between words — NSP teaches BERT to understand longer-term dependencies across sentences.

1. **What is Matthews evaluation?**

Matthew's correlation coefficient, also abbreviated as MCC was invented by Brian Matthews in 1975. MCC is a statistical tool used for model evaluation. Its job is to gauge or measure the difference between the predicted values and actual values and is equivalent to chi-square statistics for a 2 x 2 contingency table.

1. **What is Matthews Correlation Coefficient (MCC)?**

Matthew's correlation coefficient, also abbreviated as MCC was invented by Brian Matthews in 1975. MCC is a statistical tool used for model evaluation. Its job is to gauge or measure the difference between the predicted values and actual values and is equivalent to chi-square statistics for a 2 x 2 contingency table.

1. **Explain Semantic Role Labeling**

In natural language processing, semantic role labeling (also called shallow semantic parsing or slot-filling) is the process that assigns labels to words or phrases in a sentence that indicates their semantic role in the sentence, such as that of an agent, goal, or result. It serves to find the meaning of the sentence.

1. **Why Fine-tuning a BERT model takes less time than pretraining**

Finetuning pretrained language models like BERT on downstream tasks has become ubiquitous in NLP research and applied NLP. That’s in part because one can save a lot of time and money by using pretrained models. They also often serve as strong baseline models which, when finetuned, significantly outperform training models from scratch.

1. **Recognizing Textual Entailment (RTE).**

Recognizing Textual Entailment (RTE) was proposed as a unified evaluation framework to compare semantic understanding of different NLP systems. In this survey paper, we provide an overview of different approaches for evaluating and understanding the reasoning capabilities of NLP systems.

1. **Explain the decoder stack of GPT models.**

GPT model was based on Transformer architecture. It was made of decoders stacked on top of each other (12 decoders). These models were same as BERT as they were also based on Transformer architecture. ... GPT model works on a principle called autoregressive which is similar to one used in RNN.